DOCUMENT RESUME

ED 076 147 HE 004 125

AUTHOR	Graham, Margaret H.
TITLE	The Relationship between CEQ Ratings and Instructor's Rank, Class Size, and Course Level.
INSTITUTION	Illinois Univ., Urbana. Office of Instructional Resources.
REPORT NO	RR-337
PUB DATE	Feb 72
NOTE	39p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Bibliographies; *Class Size; *College Faculty; *Faculty Evaluation; *Higher Education; *Instructional Program Divisions; Research; Research Projects
IDENTIFIERS	*Illinois Course Evaluation Questionnaire

ABSTRACT

A study was conducted to determine if the tendency for faculty members of higher rank to receive the highest ratings on the Illinois Course Evaluation Questionnaire (CEQ) remained when variables such as class size and course level were taken into account. The relationship between CEQ ratings and instructor's rank, class size, and level of course was examined by means of multivariate analysis of variance (MANOVA). Dependent variables were the six subscales of the CEQ. There were no significant differences in ratings assigned by students in small, medium, and large classes, or received by teaching assistants, instructors, and assistant, associate, and full professors. Highly significant differences were found in ratings assigned in course levels. Significant size-by-level and size-by-rank interaction effects were found. Discriminant functions computed for effects found to be significant yielded information concerning the extent and direction of these significant differences. A 21-item bibliography is included. (Author/MJM)

FILMED FROM BEST AVAILABLE COPY

ED 076147

research report # 337

research report

The Relationship between CEQ Ratings and Instructor's Rank, Class Size, and Course Level

Margaret H. Graham

Measurement and Research Division
Office of Instructional Resources
University of Illinois

February, 1972

measurement and research division
office of instructional resources
university of illinois
urbana, illinois

U.S. DEPARTMENT OF HEALTH EDUCATION & WELFARE OFFICE OF EDUCATION THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

ABSTRACT

A study was conducted to determine if the tendency for faculty members of higher rank to receive the highest ratings on the Illinois Course Evaluation Questionnaire (CEQ) remained when variables such as class size and course level were taken into account. The relationship between CEQ ratings and instructor's rank, class size, and level of course was examined by means of multivariate analysis of variance (MANOVA). Dependent variables were the six subscales of the CEQ. As hypothesized there were no significant differences in ratings assigned by students in small (1-20 students), medium (21-40 students), and large (over 40 students) classes ($p < .13$), or received by teaching assistants, instructors, and assistant, associate, and full professors ($p < .13$). Highly significant differences ($p < .0001$), however, were found in ratings assigned by students in 100-, 200-, 300-, and 400- and 500-level courses. In addition, significant size by level ($p < .03$) and size by rank ($p < .02$) interaction effects were found. Discriminant functions computed for effects found to be significant yielded information concerning the extent and direction of these significant differences.

The Relationship between CEQ Ratings and Instructor's Rank, Class Size, and Course Level

by

Margaret H. Graham

During recent years, there has been an increased awareness of the need for systematic investigation into the area of teaching effectiveness, particularly at the college level. There are a number of ways in which such effectiveness can be assessed. Methods commonly employed at various institutions include classroom visitation, interviewing of students, ratings of faculty by colleagues, by students, by alumni, and by department chairman, teacher self-evaluation, and even the appraisal of student achievement. Extensive discussion of methodological considerations involved in evaluating courses and/or instructors will not be included in this report since relevant consideration of such areas can be found in sources such as Anikeef (1953), Bayley (1967), Bogue (1967), Bryant (1967), and Weaver (1960), as well as more recent works and also journals devoted specifically to the evaluation of teacher effectiveness, such as Improving College and University Teaching published by the Oregon State University Press.

One particular instrument currently being used to assess teaching effectiveness at the University of Illinois and elsewhere is the Illinois Course Evaluation Questionnaire (CEQ). The CEQ was designed to gather student reactions to a "standardized set of statements relative to certain standardized aspects of an instructional program." It has norms which "enable an instructor to adequately compare his results with results of other instructors" (Spencer and Aleamoni, 1969, p. 2). Through a procedure involving the use of factor analysis and logical grouping of items, six subscales (General Course Attitude, Method of Instruction, Course Content, Interest-Attention, Instructor, and Specific Items) were developed.

The instrument consists of 50 items (Each instructor can also include additional items of his choice.) to which students indicate their degree of agreement or disagreement. Responses to all CEQ items are based on a common four-point scale of strongly agree (SA), agree (A), disagree (D), and strongly disagree (SD). The mean item response is calculated as the numerical average of the SA, A, D, and SD responses, with a weight of 4 assigned to the most favorable response down to a weight of 1 for the least favorable response. In some cases, SA is most favorable; in others, SD occupies that position. The higher the raw score on the CEQ, the more favorable is the attitude being measured. Since some CEQ items reflect positive attitudes and some reflect negative attitudes, weights for negatively-stated items are the inversion of those for positively-stated items.

Administration of the CEQ by instructors is voluntary. CEQ results are confidential, and are presented only to the instructor who administers the instrument. Not only do such results give him insight into the manner in which he and his course are perceived by his students and thus suggest ways in which aspects of his program might be profitably altered, but they provide the Measurement and Research Division of the Office of Instructional Resources with information concerning instruction at the University as a whole, leading to constant updating of the various norms for the CEQ and the determination of characteristics such as validity and reliability of the instrument itself.

Results on the CEQ have been studied in terms of variables such as level of course (Singhal, 1968; McInnis, 1966), content of the course (Singhal, 1968), sex of the instructor (Singhal, 1968; Spencer, 1969; Singhal, 1968), rank of the instructor (Stallings and Singhal, 1969; Singhal, 1968; Spencer, 1969), instructor's severity of grading (Singhal and Stallings, 1967), class size (Spencer, 1969), grade expected by the student (Spencer, 1969; Singhal and Stallings, 1967), student status-- freshman, sophomore, junior, senior,

graduate student, other--(Spencer, 1969), and whether or not respondents have identified themselves (Spencer, 1965). Higher instructor rank, higher course level, higher student status, higher expected grade, and identification of respondent have each tended to result in a more favorable student evaluation. Of particular interest to this investigator is the tendency for faculty members of higher rank to receive the highest ratings. To what can these differences associated with rank be attributed?

STATEMENT OF PROBLEM

The purpose of this study is to attempt to determine if the tendency for faculty members of higher rank to receive the highest ratings on the CEQ remains when such variables as class size and course level are taken into account. Studies of the variable class size alone have not yielded significant results in terms of CEQ ratings. However, there might be an interaction between class size and instructor's rank. For example high rank instructors might be more favorably rated in larger classes, while low rank instructors might be more favorably rated in smaller classes. Those in smaller classes might value highly the greater opportunity for personal interaction with the instructor, thus making more positive evaluations, while those in larger classes might be favorably impressed at being taught by a person of high esteem, thus affecting their ratings. In terms of course level, perhaps high rank instructors receive higher ratings when teaching upper level courses and lower ratings when teaching lower level courses. Such a finding would be realistic since students in upper level courses, by virtue of the fact that they have remained at the University, tend to be of higher quality than those of lower level courses. Better students probably expect higher grades, and higher expected grades have been associated with higher ratings on the CEQ.

METHOD

In this study, the relationship between CEQ ratings and instructor's rank, class size, and level of course (all of which are probably correlated) will be examined by means of multivariate analysis of variance (MANOVA). Jones (1966) and Bock (1966) are among those persons who have worked extensively with this technique. Bock (1966, p. 821) pointed out that MANOVA and discriminant analysis are generalizations of methods of regression analysis, with MANOVA the generalization of the usual univariate ANOVA. MANOVA "formally includes the procedures of discriminant analysis but not its purpose" (Bock, 1966, p. 822).

Jones (1966) indicated advantages and disadvantages of using MANOVA instead of alternative techniques (e.g., correlational). He stated that, whereas the correlational techniques are often used to estimate the magnitude of association between variables of interest, MANOVA, like univariate ANOVA, is used to determine the existence of significant differences between groups or conditions. MANOVA, however, is used in studying differences among groups on many variables simultaneously. For each significant main effect, discriminant functions are computed. [See Tatsuoka's (1970) booklet for an excellent and relatively simple discussion of discriminant analysis.] The functions obtained indicate the extent and direction of the differences found to be significant. Thus a valuable feature of MANOVA is the use of discriminant analysis in interpreting significant effects and suggesting hypotheses for further study in an area.

A major problem encountered when using MANOVA concerns meeting assumptions analagous to those of univariate ANOVA. Computations required in MANOVA, as with all statistical techniques, can be made without considering these assumptions. However, if valid inferences are to be made from results

obtained, it is necessary to assume "that the within-cell residuals have a multivariate normal distribution with a common covariance matrix, and that observations on different individuals are uncorrelated" (Jones, 1966, p. 247). The latter assumption can be insured through experimental design. The former two assumptions (of normality of distribution and of homogeneity of covariance) are very difficult to prove in MANOVA, and unlike the case with univariate ANOVA, it cannot be assumed that violations of these assumptions are not critical (Jones, 1966).

Various criteria have been proposed for use in significance testing in MANOVA. Jones (1966) outlined advantages and disadvantages of the following three criteria and listed situations in which each criterion is appropriate: Wilks' maximum likelihood criterion, Hotelling's trace criterion, and Roy's largest root criterion. Bartlett's chi-square criterion has also been widely used.

Variables and Sampling

The primary sampling unit in this study is course sections. The six dependent variables are section means of the sub-scales on the CEQ as indicated in Table 1. Sub-scale means for each course section were computed so that the highest score possible was 4.00, with high scores indicating positive student attitudes toward the particular aspects of the course and instructor. The three independent variables are class size (factor A), course level (factor B), and instructor rank (factor C) for each section.

Class size, factor A, consists of three levels*:

Level 1 = small (1-20 students)

Level 2 = medium (21-40 students)

Level 3 = large (over 40 students).

Course level, factor B, consists of four levels:

Level 1 = 100-level courses

Level 2 = 200-level courses

Level 3 = 300-level courses

Level 4 = 400- and 500-level courses.

Instructor rank, factor C, consists of five levels:

Level 1 = teaching assistant

Level 2 = instructor

Level 3 = assistant professor

Level 4 = associate professor

Level 5 = full professor.

Persons of rank other than the above (e.g., visiting lecturer, visiting associate professor, etc.) are not included in this study.

---

*Class size was considered to be the number of students in each section responding to the questionnaire.

Table 1

Definition of the Six Dependent Variables

Variable Number	Subscore Title	Number of Items
1	General Course Attitude (GCA)	8
2	Method of Instruction (MI)	8
3	Course Content (CC)	8
4	Interest-Attention (I-A)	8
5	Instructor (I)	8
6	Specific Items (SI)	10
	Total	50

CEQ data for a total of 4,555 course sections taught during the period from the Fall of 1965-1966 to the Summer of 1970 at the University of Illinois and other institutions were placed on electromagnetic tape during the Spring semester of 1971. These data include the following information concerning each section: subscale means, total score mean, date (semester and year) of administration of the CEQ, number of students responding to the questionnaire, name of the institution (and campus, if applicable), course level, instructor's rank, sex of the instructor, and type of class. A listing in alphabetical order was made of the contents of this tape. This listing was scanned visually, and all sections listed which met certain criteria were check-marked, indicating eligibility for inclusion in the study. Check-marked were all sections taught on the Urbana campus of the University of Illinois by instructors holding the regular rank of teaching assistant through full professor at the end of <u>any</u> semester listed. A number of sections were

excluded solely because of missing course level and/or instructor rank. Most sections taught before the Fall semester of 1966-1967 were thus automatically excluded because the incidence of missing data was highest during the earlier semesters of CEQ administration.

No instructor was represented more than once. That is, not more than one section taught by any one instructor was included in the main study. For those instructors administering the CEQ to more than one section, the section to be included was selected randomly from among all sections of a particular instructor's courses to which the CEQ was given. Sections had been listed alphabetically by instructor's name to facilitate selecting sections of distinct professors. The 440 course sections finally included in the study, therefore, provide data concerning courses taught by 440 distinct professors.

Design

The design of the study, then, is a 3 x 4 x 5 MANOVA with six dependent variables. Analysis was conducted using the IBM 360 computer and the UMAVAC program originally developed by Finn (1966) and adapted for use at the University of Illinois by Wardrop and Bligh in 1969. The analysis was carried out directly on the means of the subscores, with discriminant functions computed for those effects which were significant.

The specific hypotheses being tested are:

(1) There are no significant differences in ratings assigned by students in small (1-20), medium (21-40), and large (over 40) classes (class size).

(2) There are no significant differences in ratings assigned by students in 100-, 200-, 300-, and 400- and 500-level courses (course level).

(3) There are no significant differences in ratings received by teaching assistants, instructors, and assistant, associate, and full professors (instructor rank).

The minimum level of significance necessary to reject each null hypothesis was set at .05.

Pilot Study

At the beginning of this investigation, a pilot study involving 205 course sections was conducted. The purposes of the pilot study were to provide the investigator with insight into the nature of the concerns encountered when actually converting data into a format required for performance of a MANOVA, particularly in connection with the UMAVAC program, to enable her to successfully run such a program, to provide her with first-hand experience in examining and interpreting the resultant output from the UMAVAC program, and to suggest refinements in general procedures of the main study. Results of the pilot study are not reported, since the main emphasis in this investigation is on the main study which is based on a greater number of course sections, with sections selected for inclusion under slightly different conditions than during the pilot phase.

RESULTS

Table 2 presents the number of sections for each rank of instructor category by class size and course level. This table shows that, in terms of class size, there was a total of 26 large, 156 medium, and 236 small course sections included. In terms of course level, there was a total of 48 400- and 500-level, 61 300-level, 102 200-level, and 237 100-level course sections. And in terms of instructor rank, there was a total of 57 sections taught by full professors, and 62, 105, 60, and 84 sections taught by asso-

ciate professors, assistant professors, instructors and teaching assistants, respectively.

Table 2

Number of Sections for Each Rank of Instructor

Category by Class Size and Course Level

Rank of Instructor	Class Size	Course Level				N
		400 - 500	300	200	100	
Full Professor	Large	0	3	1	0	4
	Medium	2	8	4	1	15
	Small	11	8	13	6	38
Associate Professor	Large	0	2	0	4	6
	Medium	5	5	10	2	22
	Small	10	2	12	10	34
Assistant Professor	Large	1	3	3	1	8
	Medium	12	7	8	38	65
	Small	6	13	21	72	112
Instructor	Large	0	0	0	2	2
	Medium	0	2	2	10	14
	Small	1	3	12	28	44
Teaching Assistant	Large	0	1	1	4	6
	Medium	0	1	6	33	40
	Small	0	3	9	26	38
N		48	61	102	237	448

There were 11 empty cells, with 7 of those cells involving the highest (400- and 500-level) level courses. Such a finding is reasonable since only 10.7 per cent (n = 48) of the sections represented are of the highest course level.

As shown in Table 2, the highest cell frequency (n = 72) involves assistant professors teaching small 100-level sections. There were 38 assistant professors and 33 teaching assistants teaching medium-sized 100-level sections. Only one large 400- and 500-level course was included.

All in all, the entries in Table 2 seem reasonable and would probably be representative of frequencies in each cell if selection had been based on all the courses on the Urbana campus rather than those sections the instructor of which chose to administer the CEQ.

The means for each cell on each of the dependent variables and the number of sections on which each mean is based are contained in Table 3. As indicated above, the highest possible sub-scale mean is 4.00, with higher scores indicating more positive student attitudes toward the course and its instructor. As noted at the bottom of this table, definition of each of the three factors and their respective levels is found on pages 5 and 6 of this report.

Table 3

Means for Each Cell on Each

of the Dependent Variables

Cell*	N**	OCA	MI	CC	I-A	I	SI
$A_1B_1C_1$	26	3.02	2.79	2.76	2.65	3.13	2.82
$A_1B_1C_2$	28	3.06	2.89	2.85	2.81	3.15	2.91
$A_1B_1C_3$	72	3.02	2.79	2.79	2.77	3.16	2.86
$A_1B_1C_4$	10	2.93	2.55	2.66	2.60	2.91	2.77
$A_1B_1C_5$	6	3.07	2.59	2.75	2.67	3.08	2.84
$A_1B_2C_1$	9	3.06	2.75	2.81	2.74	3.04	2.92
$A_1B_2C_2$	12	3.22	2.99	3.05	2.99	3.27	3.04
$A_1B_2C_3$	21	3.21	2.93	3.01	2.92	3.29	2.97
$A_1B_2C_4$	12	3.20	2.93	3.03	2.92	3.33	3.01
$A_1B_2C_5$	13	3.03	2.69	2.91	2.68	3.21	2.88
$A_1B_3C_1$	3	3.40	3.05	3.04	3.18	3.39	3.13
$A_1B_3C_2$	3	3.31	2.81	2.93	3.14	3.37	2.95
$A_1B_3C_3$	13	3.27	2.81	2.99	2.98	3.26	2.98
$A_1B_3C_4$	2	3.44	3.14	3.04	3.19	3.38	3.01
$A_1B_3C_5$	8	3.07	2.62	2.83	2.64	3.02	2.83
$A_1B_4C_2$	1	3.05	3.07	3.02	2.84	3.20	2.98
$A_1B_4C_3$	6	3.40	3.10	3.16	3.20	3.39	3.17
$A_1B_4C_4$	10	3.30	2.98	2.99	3.08	3.30	2.94
$A_1B_4C_5$	11	3.33	2.97	3.04	3.06	3.23	2.98
$A_2B_1C_1$	33	3.00	2.74	2.81	2.68	3.10	2.80
$A_2B_1C_2$	10	3.23	3.08	3.05	3.06	3.33	2.99
$A_2B_1C_3$	38	3.10	2.82	2.87	2.75	3.13	2.91

(Table 3 Cont.)

Cell*	N**	GCA	HM	CC	I-A	I	SI
$A_2B_1C_4$	2	3.13	2.96	2.82	2.82	3.16	2.90
$A_2B_1C_5$	1	3.62	2.79	3.12	3.01	3.12	3.14
$A_2B_2C_1$	6	3.10	2.99	2.92	2.80	3.30	2.99
$A_2B_2C_2$	2	3.55	3.43	3.23	3.36	3.52	3.23
$A_2B_2C_3$	8	3.20	3.02	3.07	3.02	3.35	3.05
$A_2B_2C_4$	10	2.95	2.57	2.83	2.57	3.00	2.84
$A_2B_2C_5$	4	2.68	2.65	2.76	2.45	3.19	2.87
$A_2B_3C_1$	1	3.24	2.97	2.98	3.08	3.26	2.91
$A_2B_3C_2$	2	3.25	2.41	2.98	2.76	3.17	3.04
$A_2B_3C_3$	7	3.09	2.70	2.81	2.75	3.09	2.85
$A_2B_3C_4$	5	3.25	2.62	2.88	2.89	3.16	2.89
$A_2B_3C_5$	8	3.32	3.13	2.96	3.04	3.45	3.02
$A_2B_4C_3$	12	3.22	2.84	2.98	2.88	3.23	2.96
$A_2B_4C_4$	5	3.06	2.73	2.83	2.73	3.26	2.86
$A_2B_4C_5$	2	3.12	2.69	2.86	2.69	3.13	2.78
$A_3B_1C_1$	4	3.03	2.84	2.86	2.83	3.19	2.83
$A_3B_1C_2$	2	3.27	2.92	2.98	3.00	3.05	2.94
$A_3B_1C_3$	1	3.13	2.95	2.77	2.82	3.15	2.77
$A_3B_1C_4$	4	3.05	2.68	2.82	2.68	3.15	2.76
$A_3B_2C_1$	1	2.97	2.87	2.80	2.73	3.01	2.82
$A_3B_2C_3$	3	3.28	2.97	3.13	2.95	3.16	3.03
$A_3B_2C_5$	1	3.35	2.92	3.01	2.93	3.12	3.99
$A_3B_3C_1$	1	2.85	2.29	2.80	2.41	2.91	2.90

(Table 3 cont.)

Cell*	N**	GCA	MI	CC	I-A	I	SI
$A_3B_3C_3$	3	2.88	2.20	2.67	2.39	2.69	2.74
$A_3B_3C_4$	2	3.25	2.88	3.02	2.92	3.26	2.92
$A_3B_3C_5$	3	3.06	2.62	2.67	2.74	3.05	2.83
$A_3B_4C_3$	1	2.74	2.17	2.49	2.22	2.78	2.71

*See pages 5-6 for definition of Factors A, B, and C and their respective levels.

**N = number of sections in each cell.

Tables 4, 5, and 6 present results of tests of significance for each of the three main effects, and Tables 8, 9, and 10 present results for each of the two-way interactions. The only main effect found to be significant was the level effect. In addition, the size x level and the size x rank interactions were significant.

In Table 4, the size effect is shown to be nonsignificant (Multivariate F = 1.48, d.f. = 12 and 816, $p < .13$). Thus, there were no significant differences in ratings in terms of class size. In this and other tables reporting results of tests of significance, results of Multivariate F, Univariate F, and Step-down F tests are reported. In Table 4, it is evident that none of the Univariate F results are significant, but that the value of 3.40 indicated for the Step-down F for the Interest-Attention variable is significant at the .05 level. Results of the univariate tests of significance show the effect of the individual dependent variables. Thus, for the size effect, none of the six

dependent variables taken singly produced significant differences in CEQ ratings between small, medium, and large course sections. Step-down F results indicate effects of each dependent variable when variance attributed to other dependent variables is partialed out. Significant values of Step-down F indicate the variable(s) primarily responsible for significant values of the (overall) Multivariate F.

Table 4

Tests of Significance

for the Size Effect

Dependent Variable	Univariate F (df_h=2, df_e=413)	Step-down F (df_h=2, df_e=413)	Multivariate F (df_h=12, df_e=816)
General Course Attitude	.10	.10	
Method of Instruction	1.28	2.76	
Course Content	.40	1.05	F = 1.48
Interest-Attention	.82	3.40*	
Instructor	1.89	.61	p< .13
Specific Items	.12	.94	

*p < .05

Table 5 shows that the level effect was highly significant (Multivariate F = 6.40, d.f. = 18 and 1154.4034, p < .0001). The variables primarily responsible for the significant level effect are General Course Attitude (Step-down F = 6.31, p < .001), Method of Instruction (Step-down F = 11.08, p < .001), and Course Content (Step-down F = 13.38, p < .001. The step-down F values for the Interest-Attention and Instructor variables were also significant at the .05 level, with the only variable for whom a nonsignificant value of Step-down F was found being Specific Items. There were two significant discriminant functions (Bartlett's X^2_1 = 110.6064, d.f. = 18, p < .00001; Bartlett's X^2_2 = 40.68, d.f. = 10, p < .00001) for the level effect:

$$Y_1 = -.78X_1 + 2.20X_2 - .83X_3 - .24X_4 - .60X_5 - .28X_6$$

or

Y_1 = .78 (General Course Attitude) + 2.20 (Method of Instruction) - .83 (Course Content) - .24 (Interest-Attention) - .60 (Instructor) - .28 (Specific Items).

$$Y_2 = 1.11X_1 - 1.01X_2 - 1.67X_3 + 1.40X_4 + .24X_5 - .24X_6$$

or

Y_2 = 1.11 (General Course Attitude) - 1.01 (Method of Instruction) - 1.67 (Course Content) + 1.40 (Interest-Attention) + .24 (Instructor) - .24 (Specific Items).

A plot of these functions is presented in Figure 1. The first function, Y_1, is based mostly on high scores on Method of Instruction and low scores on Course Content and General Course Attitude. On this function, 100- and 400- and 500-level course sections were very similar. The 300-level sections were much lower on this function than courses of other levels, with the 200-level sections intermediate (between 100- and 400-level and 300-level sections).

The second function, Y_2, is based primarily on high scores on General Course Attitude and Interest-Attention and low scores on Course Content. This function separates 200-level sections from all others, and tends to spread out the remaining three levels from one another, but not very far, with 300-level highest, 100-level next, then 400- and 500-level, and finally 200-level sections lowest.

Thus, for the first function for the level effect, high Method of Instruction and low Course Content and General Course Attitude ratings separate 100- and 400- and 500-level courses from courses of the other two levels, with 300-level courses lowest and 200-level courses intermediate on this function.

Table 5

Tests of Significance

for the Level Effect

Dependent Variable	Univariate F (df_h=3, df_e=413)	Step-down F (df_h=3, df_e=413)	Multivariate F (df_h=18, df_e=1154.4834)	Standardized Discriminant Coefficients 1	2
General Course Attitude	6.31***	6.31***		-.78	1.11
Method of Instruction	1.69	11.08***		2.20	-1.01
Course Content	10.04***	13.38***	F = 6.40	-.83	-1.67
Interest-Attention	4.56**	3.08*	p < .0001	-.24	1.40
Instructor	3.25*	3.22*		-.60	0.24
Specific Items	6.95***	1.09		-.28	-0.24

* $p < .05$
** $p < .01$
*** $p < .001$

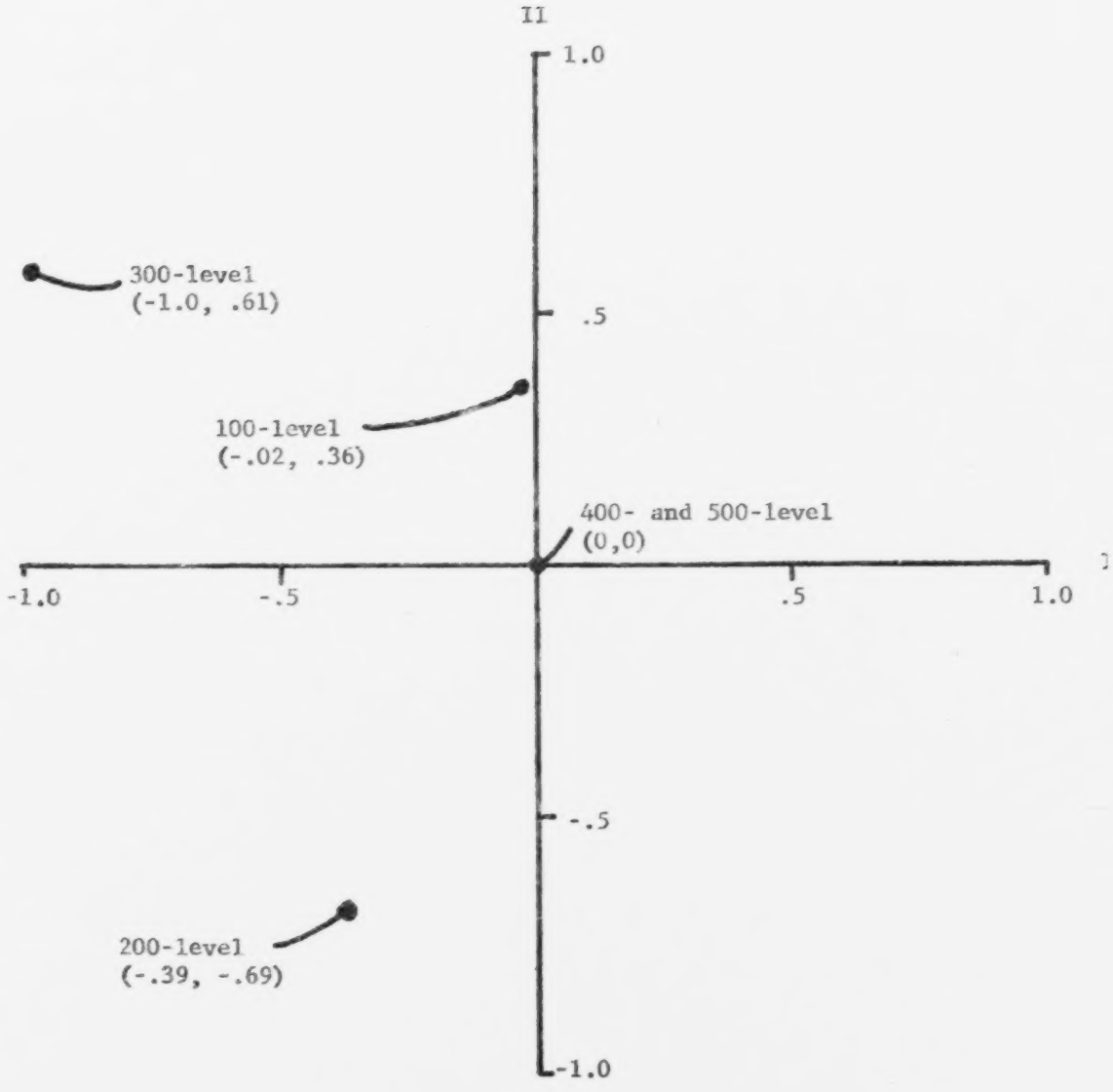

Figure 1. Centroids of the four groups based on course level in the discriminant space.

And for the second function, high General Course Attitude and Interest-Attention and low Course Content ratings separate 200-level sections from all the rest, with 300-level sections highest on this function.

Table 6 shows results of the tests of significance for the rank of instructor effect. There were no significant differences by rank (Multivariate F = 1.35, d.f. = 24 and 1424.5520, $p < .13$). However, as the univariate tests of significance showed, if the variables were examined one at a time, significant differences would have been found among Method of Instruction, Course Content, Interest-Attention, and Specific Items with Univariate F values of 2.48, 2.71, 2.75, and 3.07, respectively, all significant at the .05 level. Table 7, however, shows that there are high intercorrelations among the six dependent variables (r = .67 - .91). Thus, the four significant differences found during univariate tests are artifacts of the high intercorrelations among the CEQ subscales. When the intercorrelations of each of the four significant (during univariate analysis) scales are taken into account, in addition to their intercorrelations with the two remaining scales, the remaining differences are no longer found to be significant. Thus, although univariate tests had revealed significant rank differences, multivariate tests did not. The multivariate tests take into account the relationships among the dependent variables, thus presenting a more accurate account of events being studied.

Table 6

Tests of Significance

for the Rank Effect

Dependent Variable	Univariate F (df_h=4, df_e=413)	Step-down F (df_h=4, df_e=413)	Multivariate F (df_h=24, df_e=1424.5520)
General Course Attitude	1.23	1.23	
Method of Instruction	2.48*	2.17	F = 1.35
Course Content	2.71*	1.27	p < .13
Interest-Attention	2.75*	1.65	
Instructor	1.00	.76	
Specific Items	3.07*	1.02	

Table 7

Intercorrelations Among the Six Dependent Variables

	Dependent Variable	1	2	3	4	5	6
1	General Course Attitude	1.00					
2	Method of Instruction	.85	1.00				
3	Course Content	.86	.84	1.00			
4	Interest-Attention	.91	.89	.87	1.00		
5	Instructor	.67	.81	.72	.73	1.00	
6	Specific Items	.77	.80	.86	.78	.69	1.00

* p < .05

Tests of significance for the size x level interaction are shown in Table 8. The results were significant at the .03 level (Multivariate F = 1.49, d.f. = 36 and 1794.4136). The variables primarily responsible for the significant interaction between size and level are Course Content and Specific Items. The one significant discriminant function for this interaction effect (Bartlett's X^2 = 53.53, d.f. = 36, $p < .04$) is:

$$Y_1 = .37X_1 + .26X_2 - 2.08X_3 - .28X_4 + .03X_5 + 1.27X_6$$

or

Y_1 = .37 (General Course Attitude) + .26 (Method of Instruction) - 2.08 (Course Content) - .28 (Interest-Attention) + .03 (Instructor) + 1.27 (Specific Items).

This function, then, is based primarily on low scores on Course Content and high scores on Specific Items. The distribution of sections according to class size and course level is found in Figure 2.

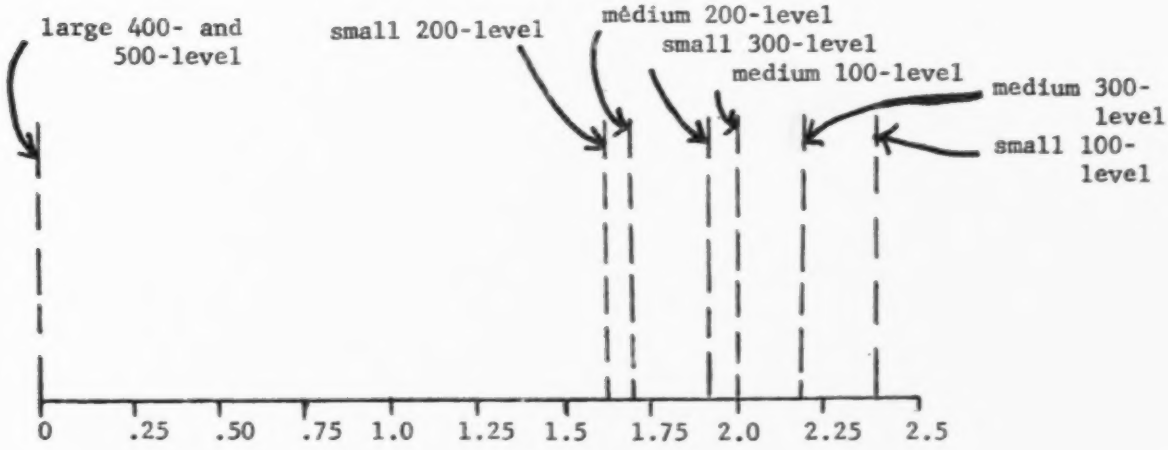

Figure 2. Canonical weights of sections according to class size and course level.

Table 8

Tests of Significance for

the Size X Level Interaction

Dependent Variable	Univariate F (df_h=6, df_e=413)	Step-down F (df_h=6, df_e=413)	Multivariate F (df_h=36, df_e=1794.4136)	Standardized Discriminant Coefficients
General Course Attitude	1.68	1.68	F = 1.49	.37
Method of Instruction	2.08	1.12	p < .03	.26
Course Content	2.82*	2.14*		-2.08
Interest-Attention	2.27*	.27		-.28
Instructor	1.11	1.16		.03
Specific Items	1.78	2.62*		1.27

* p < .05

This function, Y_1, seems to primarily separate large 400- and 500-level sections from the rest of the group. Large 400- and 500-level courses are quite separate from the rest of the sections and lowest on this function. In addition, it seems to separate 200-level sections from 100- and 300-level sections. The medium-sized sections cannot be completely separated from the small sections because the medium classes are embedded within the small group.

Group centroids on Y_1 are given below in Table 9. The numbers in parentheses indicate the number of sections on which each centroid is based. It should be noted that there was only one large 400- and 500-level section, a fact which makes interpretation of this function rather difficult. The difference between the centroids of small and medium sections for each level is given so that interaction effects can be viewed more clearly.

Table 9

Group Centroids

Course Level	Class Size		
	Small	Medium	Small - Medium
100	$2.42_{(142)}$	$2.01_{(84)}$	+.41
200	$1.61_{(67)}$	$1.71_{(30)}$	-.10
300	$1.92_{(29)}$	$2.18_{(23)}$	-.26

These results show that at the 100-level, instructors of small classes received higher ratings on the CEQ on this function. At the 200-level, there was not much difference between ratings received by instructors of small and medium classes. At the 300-level, however, instructors of medium-sized classes received higher ratings than did instructors of small classes.

Figure 3 shows a plot of the centroids for small and medium classes at each of the three course levels. In this figure it can be seen that the lines for small and medium classes cross over indicating significant interaction between class size and course level.

Thus, the significant interaction between class size and course level shows that small classes at the 100-level are rated higher on Specific Items and lower on Course Content than are medium classes at the same level. Medium classes at the 300-level are rated higher on Specific Items and lower on Course Content than are small classes at that level. There seems to be no appreciable difference between ratings assigned by small medium classes at the 200-level.

Table 10 shows that the size x rank interaction was significant (Multivariate $F = 1.47$, d.f. = 48 and 2011.5935, $p < .02$). The variables primarily responsible for the significant interaction are Method of Instruction (Step-down $F = 3.45$, $p < .05$) and Specific Items (Step-down $F = 3.45$, $p < .001$). It should be noted that although the values of Step-down F for these two variables were significant, significant values for Univariate F had not been found for any of the six dependent variables. The one significant discriminant function for this interaction effect (Bartlett's $X^2 = 70.35$, d.f. = 48, $p < .02$) is:

$$Y_1 = -.03X_1 + .28X_2 - 2.13X_3 + .55X_4 - .04X_5 + 1.67X_6$$

or

Y_1 = -.03 (General Course Attitude) + .28 (Method of Instruction) - 2.13 (Course Content) + .55 (Interest-Attention) - .04 (Instructor) + 1.67 (Specific Items).

This function, then is based primarily on low scores on Course Content and high scores on Specific Items. The distribution of sections according to class size and instructor rank on this function is shown in Figure 4, with

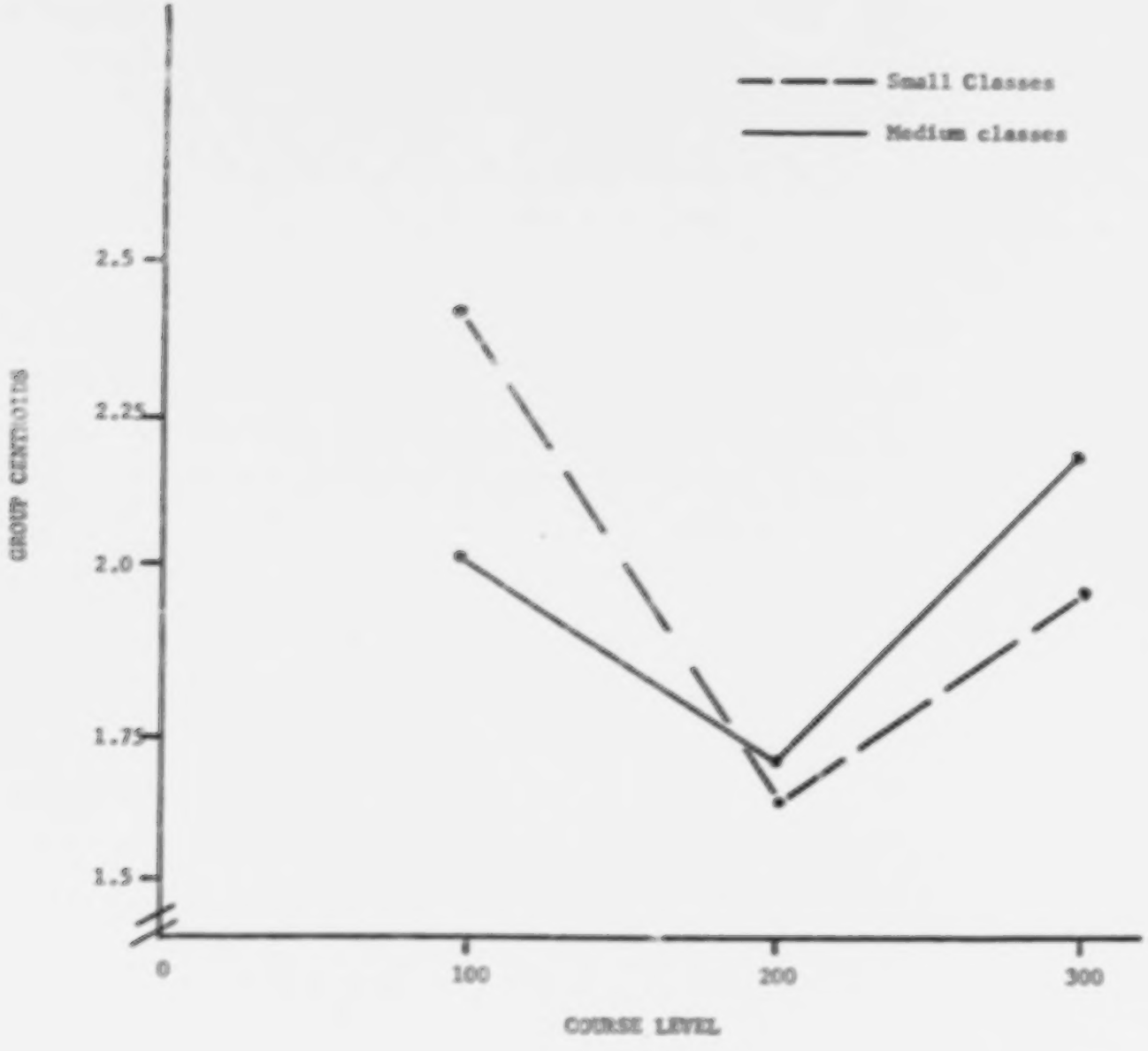

Figure 1. Group centroids on the significant discriminant function for the interaction between class size and course level.

Table 10

Tests of Significance for the

Size x Rank Interaction

Dependent Variable	Univariate F (df_h=8, df_e=413)	Step-down F (df_h=8, df_e=413)	Multivariate F (df_h=48, df_e=2011.5935)	Standardized Discriminant Coefficients
General Course Attitude	.48	.48	F = 1.47	-.03
Method of Instruction	1.12	2.00*	p < .02	.28
Course Content	.54	1.25		-2.13
Interest-Attention	.75	1.05		.55
Instructor	1.41	.67		-.04
Specific Items	1.39	3.45**		1.67

* p <.05
** p <.01

size indicated by L, M, and S, for large, medium, and small, respectively, and rank indicated by 1, 2, 3, 4, and 5, for teaching assistant, instructor, assistant professor, associate professor, and full professor, respectively:

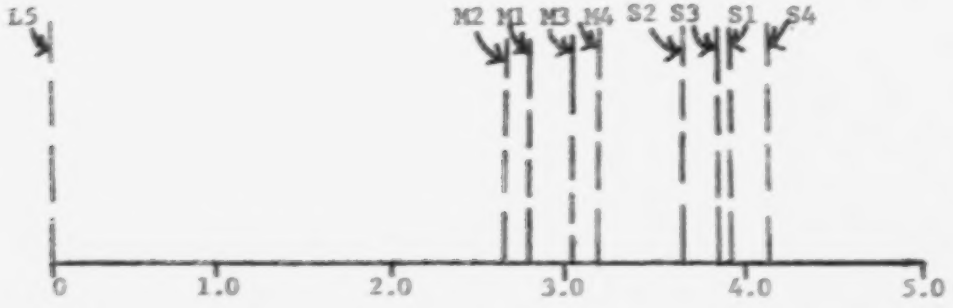

Figure 4. Canonical weights of instructor rank and section size.

This function seems to clearly separate large sections taught by full professors from all the remaining sections. In addition, medium sections are separated from small sections. Large sections taught by full professors were lowest, medium-sized sections were next (intermediate), and small sections were highest, on this function. Of those teaching medium-sized sections, instructors and teaching assistants were similar on this function, as were assistant and associate professors. Of those teaching small sections assistant professors and teaching assistants were quite similar on this function.

Group centroids on Y_1 are given in Table 11, again with numbers in parentheses indicating the number of sections on which each centroid is based. There were four full professors teaching large sections.

Differences between centroids of small and medium-sized groups for each instructor rank are also included.

Table 11

Group Centroids

Instructor Rank	Class Size		
	Small	Medium	Small - Medium
Teaching Assistant	$3.95_{(38)}$	$2.76_{(40)}$	+1.19
Instructor	$3.78_{(44)}$	$2.71_{(14)}$	+1.07
Assistant Professor	$3.93_{(112)}$	$3.04_{(65)}$	+ .89
Associate Professor	$4.18_{(34)}$	$3.24_{(22)}$	+ .94

These results show that for each rank of instructor level, instructors of small classes received higher ratings on function Y_1. That is, for each rank of instructor, lower ratings on Course Content and higher ratings on Specific Items were assigned by small than by medium classes. In addition, generally it seems that the lower the rank of instructor, the more favorable were ratings assigned by small as compared to medium sections.

A plot of the centroids for small and medium sections at each of the four rank of instructor levels is presented in Figure 5. This figure shows that for each rank level, instructors of small classes received higher ratings on this particular function.

Table 12 indicates that the level x rank interaction was not significant (Multivariate F = 1.31, d.f. = 66 and 2188.6018, $p < .06$). There were no significant values of Univariate F for this interaction, but values of Step-down F for two of the variables (Method of Instruction--Step-down F = 2.22

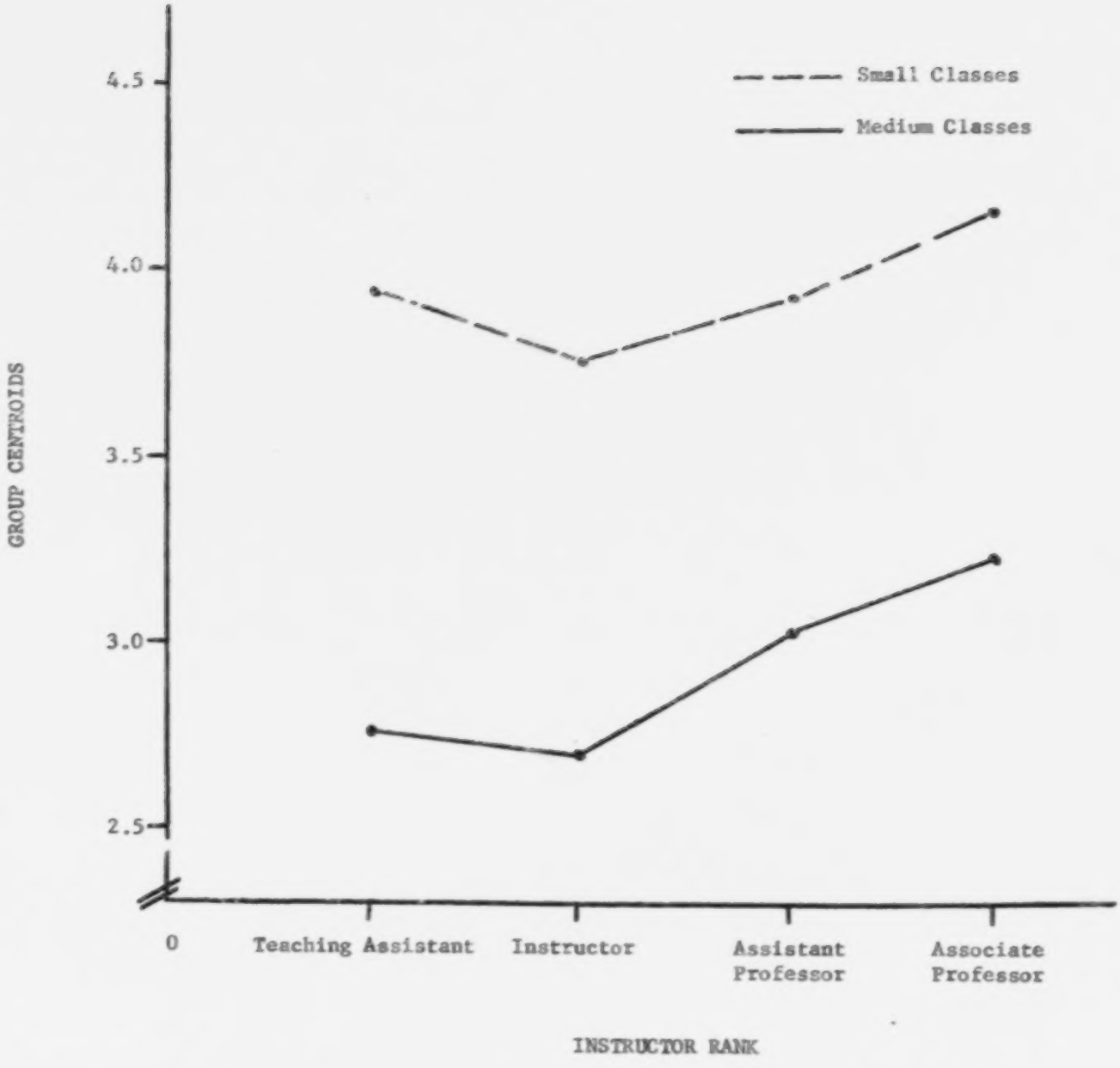

Figure 5. Group centroids on the significant discriminant function for the interaction between class size and instructor rank (main study).

Specific Items--Step-down F = 1.89) were significant at the .05 level. Therefore, if the value of Multivariate F had been significant; such significance would have been attributed primarily to Method of Instruction and Specific Items, the two variables which probably would have received the most weight in the discriminant function(s) found to be significant for this effect.

Table 12

Tests of Significance for

the Level X Rank Interaction

Dependent Variable	Univariate F (df_h=11, df_e=413)	Step-down F (df_h=11, df_e=413)	Multivariate F (df_h=66, df_e=2188.6018)
General Course Attitude	.95	.95	
Method of Instruction	1.00	2.22*	F = 1.31
Course Content	.80	.88	p <.06
Interest-Attention	.81	.58	
Instructor	.78	1.33	
Specific Items	.74	1.89*	

* p <.05

SUMMARY AND CONCLUSIONS

The null hypotheses for the size and rank main effects were supported in the present study. There were, as hypothesized, no significant differences in ratings assigned by students in small (1-20), medium (21-40), and large (over 40) classes, (size alone), or received by teaching assistants, instructors, and assistant, associate, and full professors (rank alone). The hypothesis for the level effect, however, was rejected, since there were highly significant differences in ratings assigned by students in 100-, 200-, 300-, and 400- and 500-level courses. Two significant discriminant functions were computed for the level effect, pointing out the extent and direction of differences among the groups in terms of course level. The first function was based primarily on high scores on the Method of Instruction and low scores on the Course Content and General Course Attitude subscales. This function primarily separated 100- and 400- and 500-level courses which were highest, 200-level courses which were intermediate, and 300-level courses which were lowest on the function. Thus, for the first function for the level effect, 100- and 400- and 500-level courses were rated highest on the Method of Instruction and lowest on the Course Content and General Course Attitude subscales, and 300-level courses were rated lowest on the Method of Instruction and highest on the Course Content and General Course Attitude subscales. The second significant discriminant function for the level effect was based primarily on high scores on General Course Attitude and Interest-Attention and low scores on Course Content. This function separated 200-level sections from all the rest of the sections, with 300-level courses highest on the function.

Of the interaction effects, the size x level and the size x rank effects were significant, whereas the level x rank interaction was not. The one significant discriminant function for the size x level interaction primarily separated large 400- and 500-level sections (lowest on this function), as well as 200-level sections (intermediate on this function) from the rest of the groups, with medium-sized groups embedded within those of small size. It was found that at the 100-level small classes were rated higher on the Specific Items and lower on the Course Content subscales than were medium-sized classes; at the 200-level, there was no appreciable difference between small and medium-sized sections; and at the 300-level, medium-sized classes were rated higher on the Specific Items and lower on the Course Content subscale than were small classes.

The one significant discriminant function for the size x rank interaction was based primarily on low Course Content and high Specific Items subscale scores. For each rank, small classes were rated higher than medium-sized classes on this function, and the lower the instructor's rank (with the exception of assistant and associate professors which were found to be similar) the more favorable were ratings assigned by small as compared to medium-sized sections.

The fact that the multivariate F test for the rank effect did not reveal statistical significance is of great practical significance in this investigation. Univariate tests conducted in this and in other (bivariate) studies involving the CEQ have shown significant differences in terms of rank, with higher ranking instructors tending to receive higher ratings on the CEQ, but multivariate tests did not. This discrepancy was attributed to the high inter-correlations among the CEQ subscales. The multivariate tests used herein took into account the relationships among the dependent

variables, thus presenting a more accurate account of the relationship between instructor rank and CEQ ratings.

It should be noted that the results reported in this investigation are based upon that group of distinct instructors at the Urbana campus of the University of Illinois for whom complete and relevant CEQ data were available. Because of the voluntary nature of administration of the CEQ, generalizations concerning teaching effectiveness cannot be made concerning the campus as a whole. Also, since only the Urbana campus was included in the study, generalizations cannot be made concerning relationships between CEQ ratings and instructor's rank, class size, and level of course at other institutions or campuses of this institution. In addition, results obtained must be regarded with caution because of the large number of cases involved. That is, significant results obtained herein may possibly be due largely to the sheer magnitude of the degrees of freedom involved. Another possible limitation of this study concerns the classification of sections into class size. The classification of class size into the three particular categories used was based on logical analysis of the data of the pilot study. This classification, like any, causes loss of data which may have affected the results of the study. The points of trichotomization selected could have greatly affected the conclusions reached. Another investigator might have chosen a different means of classifying sections according to size. Thus, a suggestion for further study is investigation of the effect of different classification schemes for class size, including classification into other than three levels.

Further investigation into reasons for the actual results obtained should also be made. As indicated, univariate analyses have tended to show differences in CEQ ratings in terms of instructor rank. However, as pointed

out by persons such as Cattell (1966), French (1970), Fruchter (1969), Jones (1966), Roseboom (1966), and Tatsuoka (1969), human behavior is highly complex, and therefore, the multivariate approach is preferable to conducting a series of univariate analyses of bivariate data in order to determine relationships in certain areas of interest. This study has been conducted as an attempt to show that the multivariate approach is indeed, as pointed out by Cattell (1966), more economical, more consistent, and more comprehensive than is the utilization of data from a series of bivariate studies in leading to the formulation of more adequate and representative conceptions of the state of events in a particular area. In this study, an attempt has been made to show that simultaneous multivariate design does tend to eliminate the problem of statistical dependence among variables which "upsets the significance levels" when using a "series of univariate tests" (Tatsuoka, 1969). As pointed out by Tatsuoka(1969) and others, the multivariate approach is extremely powerful when applied correctly. MANOVA in particular, is useful in detecting the existence of significance differences among groups on many variables simultaneously, and in yielding information through the use of discriminant functions computed for effects found to be significant concerning the extent and direction of those significant differences.

REFERENCES

Anikeef, A. M. Factors affecting student evaluation of college faculty members. Journal of Applied Psychology, 1953, 37, 458-460.

Bayley, D. H. Making college teaching a profession. Improving College and University Teaching, 1967, 15, 115-119.

Bock, R. D. Contributions of multivariate experimental designs to educational research. In R. B. Cattell (ed.), Handbook of multivariate experimental psychology. Chicago: Rand McNally, 1966, 820-840.

Bogue, E. G. Student appraisal of teaching effectiveness in higher education: Summary of the literature. Educational Quest, 1967, 11, 6-10.

Bryant, P. T. By their fruits ye shall know them. Journal of Higher Education, 1967, 38, 326-330.

Cattell, R. B. Guest Editorial: Multivariate behavioral research and the integrative challenge. Multivariate Behavioral Research, 1966, 1, 4-23.

Finn, J. D. Univariate and multivariate analysis of variance and covariance: A program for the IBM 7094 computer, 1966. Adapted for use at the University of Illinois by J. L. Wardrop and T. J. Bligh, University of Illinois, Urbana, 1969.

French, J. W. Guest Editorial: Toward broader applications of multivariate methods. Multivariate Behavioral Research, 1970, 5, 3-4.

Fruchter, B. Guest Editorial: The scope of multivariate research, Multivariate Behavioral Research, 1969, 4, 3-4.

Jones, L. V. Analysis of variance in its multivariate developments. In R. B. Cattell (ed.), Handbook of multivariate experimental psychology. Chicago: Rand McNally, 1966, 244-266.

McInnis, T. Some methodological considerations and a report of some research findings concerning course and/or teacher evaluations by students. University of Illinois: Measurement and Research Division, Research Report No. 231, 1966.

Rozeboom, W. W. Foundations of the theory of prediction. Homewood, Illinois: Dorsey Press, 1966, 98-103.

Singhal, S. Illinois Course Evaluation Questionnaire items by rank of instructor, sex of the instructor, and sex of the student. University of Illinois: Measurement and Research, Research Report No. 282, 1968.

Singhal, S. and Stallings, W. M. A study of the relationships between course evaluations by students and severity of grading by instructors in freshman rhetoric at the University of Illinois. University of Illinois: Measurement and Research Division, Research Report No. 252, 1967.

Spencer, R. E. Course Evaluation Questionnaire anonymous vs. identified student responses. University of Illinois: Measurement and Research Division, Research Report No. 202, 1965.

Spencer, R. E. Some dimensions of the Illinois Course Evaluation Questionnaire. University of Illinois: Measurement and Research Division, Research Report No. 303, 1969.

Spencer, R. E. & Aleamoni, L. M. The Illinois Course Evaluation Questionnaire: A description of its development and a report of some of its results. University of Illinois: Measurement and Research Division, Research Report No. 292, 1969.

Stallings, W. M. and Singhal, S. Some observations on the relationships between research productivity and student evaluations of courses and teaching. University of Illinois: Measurement and Research Division, Research Report No. 274, 1969.

Tatsuoka, M. M. Multivariate Analysis, In Ch. 10, "Statistics," Review of Educational Research, December 1969, 39, 7390743.

Tatsuoka, M. M. Discriminant Analysis: The Study of Group Differences. Champaign, Illinois: The Institute for Personality and Ability Testing, 1970.

Weaver, C. M. Instructor rating by college students. Journal of Educational Psychology, 1960, 51, 21-25.